POLI176 Final Project (SP22)

Christian Bankard, Beibei Du, Hugo Lopez (ranked in alphabetical order)

```
# Our research question: What effects did the Vietnam War have on American domestic issues?
# We predict the most discussed themes in the state of the Union with respect to Vietnam topics: drugs,
# We are going to use both LDA and STM for the Topic Modelling.
#Load package libraries that needed for our final project
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tokenizers) # tokenize the data
library(quanteda) # Use this for LDA
```

```
## Package version: 3.0.0
## Unicode version: 10.0
## ICU version: 61.1
## Parallel computing: 8 of 8 threads used.
## See https://quanteda.io for tutorials and examples.
```

```
library(quanteda.textplots)
library(stm) # Structure Topic Modelling
```

```
## stm v1.3.6 successfully loaded. See ?stm for help.
##  Papers, resources, and other materials at structuraltopicmodel.com
```

```
library(seededlda)
```

```
##
## Attaching package: 'seededlda'
## The following object is masked from 'package:stats':
##
##     terms
```

```
#Set working directory, need to change this line of code
setwd("~/Desktop/POLI176:DSC161 Text as Data/Final Project")
#Load data for speeches
metadata <- read_csv("SOTU_WithText.csv")
```

```
## Rows: 236 Columns: 5
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): president, party, sotu_type, text
## dbl (1): year
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Since our focus is on Vietnam War, we need to limit the year range
# Find out the `year`range of Vietnam War
typeof(metadata$year)
```

```
## [1] "double"
```

```
class(metadata$year) # numeric
```

```
## [1] "numeric"
```

```
# select the correct years and store it into the war df
war <- metadata[which((metadata$year>=1955) & (metadata$year<=1975)),c(colnames(metadata))] # from 1 No
# war is a tibble of size 24*5

# Preprocessing of the  data and make it corpus
corpus_sotu <- corpus(war, text_field = "text") # make it into a text corpus
corpus_sotu # take a look at it
```

```
## Corpus consisting of 24 documents and 4 docvars.
## text1 :
## "Mr. President, Mr. Speaker, Members of the Congress:  First,..."
##
## text2 :
## " [Recorded on film and tape and broadcast the same day]  My ..."
##
## text3 :
## " [Read before a joint session by a clerk of the House of Rep..."
##
## text4 :
## "To the Congress of the United States:  I appear before the C..."
##
## text5 :
## "Mr. President, Mr. Speaker, Members of the 85th Congress:  I..."
##
## text6 :
## " [Delivered in person before a joint session]  Mr. President..."
##
## [ reached max_ndoc ... 18 more documents ]
```

```
#Some common pre-processing, remove the punctuation and numbers
toks <- tokens(corpus_sotu, remove_punct = TRUE, remove_numbers=TRUE) # remove punctuations and numbers
toks <- tokens_wordstem(toks) # tokenize into the wordstem
toks <- tokens_select(toks,  stopwords("en"), selection = "remove") # remove the stopwords in English
dfm <- dfm(toks) # make it the document feature matrix
dfm
```

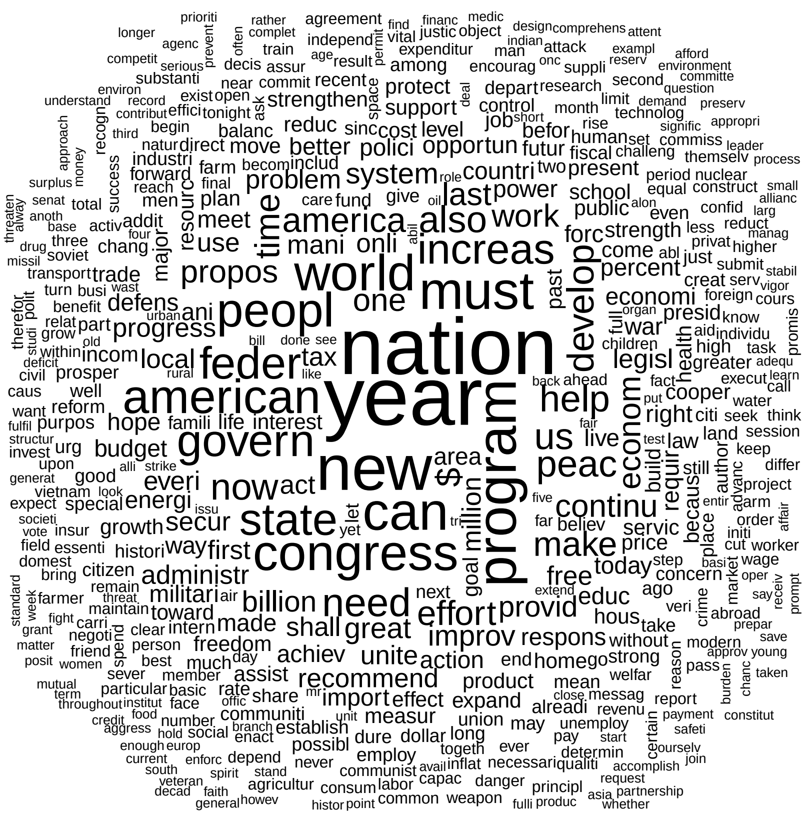
```
## Document-feature matrix of: 24 documents, 5,628 features (79.97% sparse) and 4 docvars.
##        features
## docs    mr presid speaker member congress first extend cordial greet 84th
##   text1  2      3       1      2       24    13      5       1     1    2
```

```
##   text2  0      0      0     0       6    5     0      0     0    0
##   text3  0      0      0     1      29    2     5      0     0    0
##   text4  0      0      0     1      18    7     1      0     0    0
##   text5  2      2      1     3      10    7     2      0     1    0
##   text6  2      3      1     3      16    2     2      0     0    0
## [ reached max_ndoc ... 18 more documents, reached max_nfeat ... 5,618 more features ]
```

```r
#Create a document feature matrix (dfm) and trim it with words appeared at least 5%
#toks <- corpus_sotu %>%
#  tokens()
#dfm <- dfm(toks)
dfm_trimmed <- dfm_trim(dfm, min_docfreq = 0.05, docfreq_type = "prop") # trim it and remove the words
dfm_trimmed # the cleaned up document matrix
```

```
## Document-feature matrix of: 24 documents, 3,275 features (68.58% sparse) and 4 docvars.
##        features
## docs    mr presid speaker member congress first extend greet shall much
##   text1  2      3       1      2       24    13      5     1    29    2
##   text2  0      0       0      0        6     5      0     0     0    1
##   text3  0      0       0      1       29     2      5     0    19   10
##   text4  0      0       0      1       18     7      1     0     8    2
##   text5  2      2       1      3       10     7      2     1     5    2
##   text6  2      3       1      3       16     2      2     0    11    5
## [ reached max_ndoc ... 18 more documents, reached max_nfeat ... 3,265 more features ]
```

```r
# the word cloud for documents from the 20 years of Vietnam War
textplot_wordcloud(dfm_trimmed, col="black")
```




```r
############ LDA ############
```

```r
#Run LDA using quanteda
lda <- textmodel_lda(dfm_trimmed, k = 6)
#Most likely term for each topic
lda.terms <- terms(lda, 6) # check the top words for each topic
lda.terms
```

```
##      topic1   topic2    topic3    topic4      topic5    topic6
## [1,] "world"  "program" "also"    "nation"    "$"       "american"
## [2,] "nation" "govern"  "new"     "free"      "must"    "year"
## [3,] "must"   "continu" "increas" "freedom"   "year"    "can"
## [4,] "peac"   "year"    "propos"  "help"      "billion" "peopl"
## [5,] "secur"  "feder"   "feder"   "communist" "presid"  "new"
## [6,] "econom" "state"   "year"    "effort"    "work"    "nation"
```

```r
#Topical content matrix
mu <- lda$phi
dim(mu) #6 topics
```

```
## [1]    6 3275
```

```r
#Most representative words in Topic 1-10 using a for loop
for (i in 1:6){
  print(mu[i,][order(mu[i,], decreasing=T)][1:6])
  print("############################################################")
}
```

```
##      world     nation       must       peac      secur     econom
## 0.01849104 0.01753655 0.01710269 0.01528049 0.01467309 0.01371860
## [1] "############################################################"
##    program     govern    continu       year      feder      state
## 0.01794012 0.01421573 0.01387194 0.01215299 0.01112162 0.01094972
## [1] "############################################################"
##       also        new    increas     propos      feder       year
## 0.02050023 0.01929057 0.01793859 0.01402498 0.01295763 0.01210375
## [1] "############################################################"
##      nation        free     freedom        help   communist      effort
## 0.023675801 0.012388658 0.009396886 0.009124907 0.008988917 0.008716938
## [1] "############################################################"
##          $       must       year    billion     presid       work
## 0.01797683 0.01649012 0.01339280 0.01178220 0.01116273 0.01103884
## [1] "############################################################"
##   american       year        can      peopl        new     nation
## 0.02145858 0.01979733 0.01937015 0.01732919 0.01704440 0.01552555
## [1] "############################################################"
```

```r
#Topical prevalence matrix
pi <- lda$theta
dim(pi) #24 10
```

```
## [1] 24  6
```

```r
#Most representative documents in Topic 1-10
for (i in 1:6){
  print(war[order(pi[i,],decreasing=T),])
  print("############################################################")
}
```

```
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 2 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## 3 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 4 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 5 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 6 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 2 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## 3 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 4 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## 5 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 6 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 2 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## 3 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 4 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 5 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 6 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## 2 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 3 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 4 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 5 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 6 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## 2 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 3 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 4 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 5 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## 6 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
```

```
## 2 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 3 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 4 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 5 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 6 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## [1] "############################################################"
```

```
############ STM ############
# Get the `clusterFightinWords`function from Discussion3.R; We will only define this function once
clusterFightinWords <- function(dfm, clust.vect, alpha.0=100) {
  overall.terms <- colSums(dfm)
  # n and n_k in Monroe et al.
  n <- sum(overall.terms)
  # alpha_{kw} in Monroe et al.
  prior.terms <- overall.terms / n * alpha.0
  # y_{kw}(i) in Monroe et al.
  cluster.terms <- colSums(dfm[clust.vect, ])
  # n_k(i) in Monroe et al.
  cluster.n <- sum(cluster.terms)
  cluster.term.odds <-
    (cluster.terms + prior.terms) /
    (cluster.n + alpha.0 - cluster.terms - prior.terms)
  overall.term.odds <-
    (overall.terms + prior.terms) /
    (n + alpha.0 - overall.terms - prior.terms)
  log.odds <- log(cluster.term.odds) - log(overall.term.odds)
  variance <- 1/(cluster.terms + prior.terms) + 1/(overall.terms + prior.terms)
  output <- log.odds / sqrt(variance)
  names(output) <- colnames(dfm)
  return(output)
}
# Find words that are distinctive between newsletters written by Democrats and Republicans
terms <- clusterFightinWords(dfm_trimmed,war$sotu_type == "speech")
sort(terms, decreasing=T)[1:6] #nation,people
```

```
##    peopl  america      let     free       us     peac
## 3.362773 3.070883 2.798004 2.750288 2.710062 2.643997
```

```
terms <- clusterFightinWords(dfm_trimmed,war$sotu_type == "written")
sort(terms, decreasing=T)[1:6] #federal, administration, veterans, education
```

```
##        also   administr       feder     program environment     veteran
##    4.784610    4.246175    4.128029    3.567095    3.299470    3.283223
```

```
# Structural Topic Model
#STM
#library(tm)
temp<-textProcessor(documents=war$text,metadata=war)
```

```
## Building corpus...
## Converting to Lower Case...
## Removing punctuation...
## Removing stopwords...
## Removing numbers...
## Stemming...
## Creating Output...
```

```r
out <- prepDocuments(temp$documents, temp$vocab, temp$meta)
```

```
## Removing 3101 of 6301 terms (3101 of 27073 tokens) due to frequency
## Your corpus now has 24 documents, 3200 terms and 23972 tokens.
```

```r
# Run the model
# Should we change the value of K
model.stm <- stm(out$documents, out$vocab, K = 6, prevalence = ~ party,
                 data = out$meta, max.em.its = 10)
```

```
## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      ......
##   Recovering initialization...
##      ..............................
## Initialization complete.
## .......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -6.918)
## .......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 2 (approx. per word bound = -6.829, relative change = 1.284e-02)
## .......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 3 (approx. per word bound = -6.797, relative change = 4.635e-03)
## .......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 4 (approx. per word bound = -6.787, relative change = 1.490e-03)
## .......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 5 (approx. per word bound = -6.782, relative change = 7.502e-04)
## Topic 1: program, will, nation, year, must
##  Topic 2: will, year, new, nation, can
##  Topic 3: nation, will, must, new, year
##  Topic 4: will, year, feder, new, program
##  Topic 5: will, year, nation, congress, can
##  Topic 6: nation, will, world, must, year
## .......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 6 (approx. per word bound = -6.780, relative change = 3.306e-04)
## .......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 7 (approx. per word bound = -6.778, relative change = 1.833e-04)
## .......................
## Completed E-Step (0 seconds).
## Completed M-Step.
```

```
## Completing Iteration 8 (approx. per word bound = -6.778, relative change = 1.273e-04)
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 9 (approx. per word bound = -6.777, relative change = 1.009e-04)
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Model Terminated Before Convergence Reached
```

```
model.stm <- stm(out$documents, out$vocab, K = 6, prevalence = ~ party,
                 data = out$meta)
```

```
## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      ......
##   Recovering initialization...
##      ..............................
## Initialization complete.
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -6.918)
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 2 (approx. per word bound = -6.829, relative change = 1.284e-02)
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 3 (approx. per word bound = -6.797, relative change = 4.635e-03)
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 4 (approx. per word bound = -6.787, relative change = 1.490e-03)
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 5 (approx. per word bound = -6.782, relative change = 7.502e-04)
## Topic 1: program, will, nation, year, must
##  Topic 2: will, year, new, nation, can
##  Topic 3: nation, will, must, new, year
##  Topic 4: will, year, feder, new, program
##  Topic 5: will, year, nation, congress, can
##  Topic 6: nation, will, world, must, year
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 6 (approx. per word bound = -6.780, relative change = 3.306e-04)
## ......................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 7 (approx. per word bound = -6.778, relative change = 1.833e-04)
## ......................
```

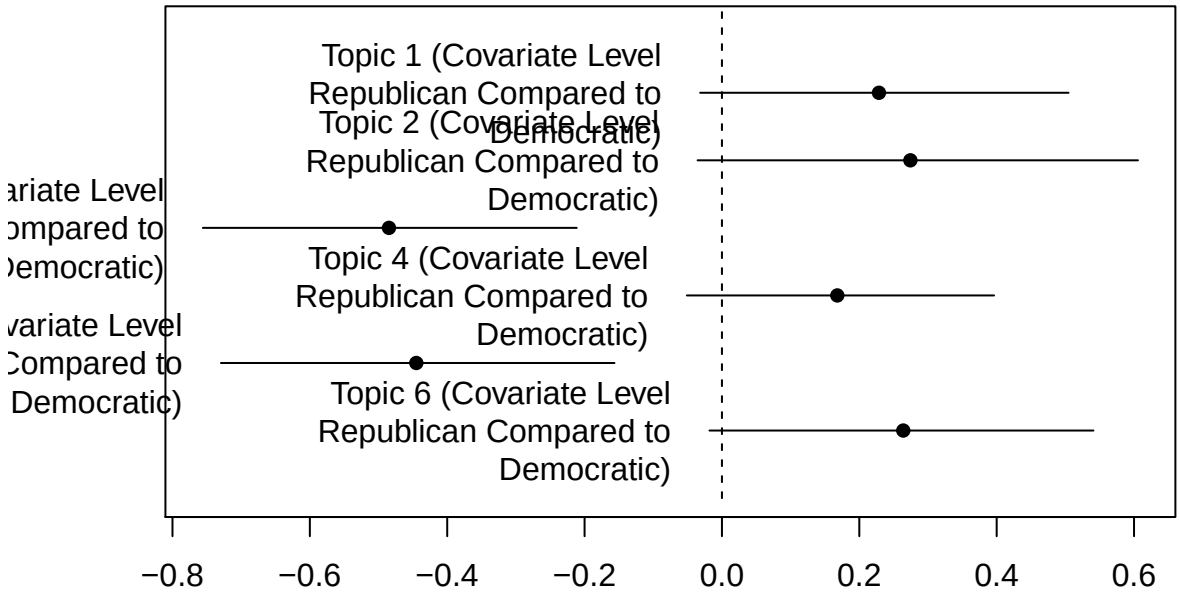
```
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 8 (approx. per word bound = -6.778, relative change = 1.273e-04)
## .....................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 9 (approx. per word bound = -6.777, relative change = 1.009e-04)
## .....................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 10 (approx. per word bound = -6.776, relative change = 8.710e-05)
## Topic 1: nation, program, will, year, must
##  Topic 2: will, year, new, nation, can
##  Topic 3: nation, will, must, new, year
##  Topic 4: will, year, new, program, feder
##  Topic 5: will, year, nation, can, congress
##  Topic 6: nation, world, will, must, year
## .....................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 11 (approx. per word bound = -6.776, relative change = 6.381e-05)
## .....................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 12 (approx. per word bound = -6.776, relative change = 2.763e-05)
## .....................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 13 (approx. per word bound = -6.775, relative change = 1.904e-05)
## .....................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 14 (approx. per word bound = -6.775, relative change = 1.420e-05)
## .....................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 15 (approx. per word bound = -6.775, relative change = 1.066e-05)
## Topic 1: nation, program, will, year, must
##  Topic 2: will, year, new, nation, can
##  Topic 3: nation, will, must, new, year
##  Topic 4: will, year, new, program, feder
##  Topic 5: will, year, nation, can, congress
##  Topic 6: nation, world, will, must, year
## .....................
## Completed E-Step (0 seconds).
## Completed M-Step.
## Model Converged
```

```
labelTopics(model.stm)
```

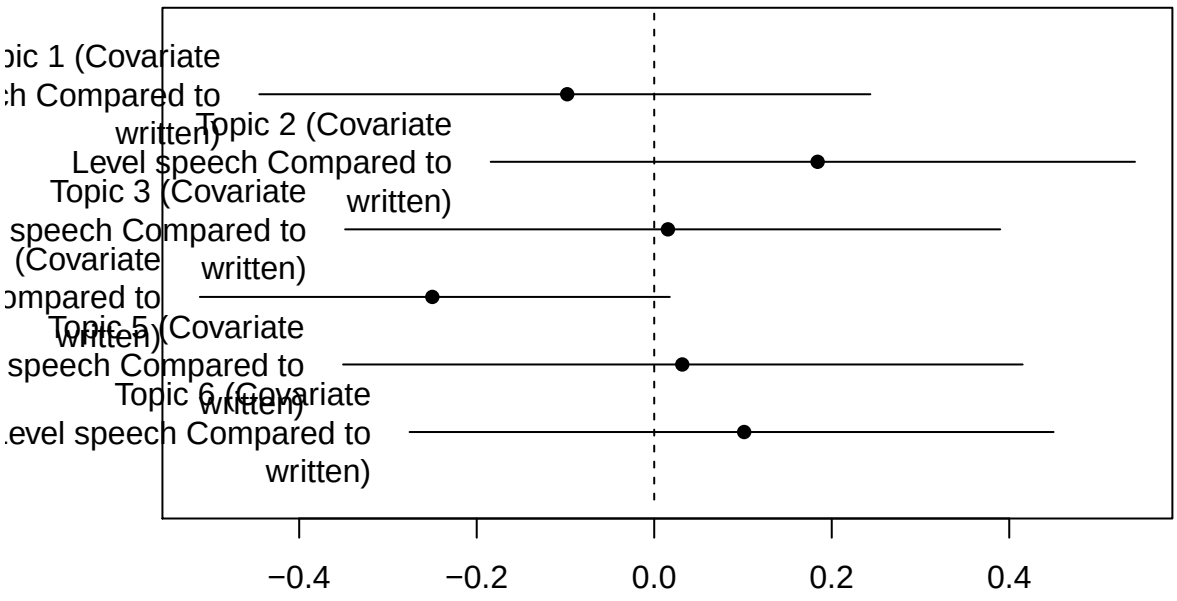
```
## Topic 1 Top Words:
##       Highest Prob: nation, program, will, year, must, govern, feder
##       FREX: citizen, immigr, activ, agricultur, vigor, recommend, atom
##       Lift: acreag, contributori, film, imprison, itali, old-ag, overdu
##       Score: film, mention, old-ag, statehood, survivor, subvers, suffrag
```

```
## Topic 2 Top Words:
##       Highest Prob: will, year, new, nation, can, america, american
##       FREX: america, let, chanc, open, centuri, generat, place
##       Lift: disagre, fed, freshmen, loser, monstrous, outragean, partisanship
##       Score: disagre, tonight, seventi, chamber, colleagu, reform, agenda
## Topic 3 Top Words:
##       Highest Prob: nation, will, must, new, year, world, can
##       FREX: allianc, instead, lack, west, atlant, recess, vote
##       Lift: -hour, buri, chao, full-tim, insid, overshadow, pile
##       Score: full-tim, viet-nam, choic, gold, cold, anew, instead
## Topic 4 Top Words:
##       Highest Prob: will, year, new, program, feder, also, increas
##       FREX: environment, energi, oil, also, reform, drug, indian
##       Lift: disloc, petroleum, buse, pleas, automat, automot, decontrol
##       Score: petroleum, environment, bicentenni, oil, reform, televis, minor
## Topic 5 Top Words:
##       Highest Prob: will, year, nation, can, congress, american, peopl
##       FREX: tonight, vietnam, tri, think, presid, south, want
##       Lift: -tax, afternoon, ambassador, battleth, belov, chairman, consular
##       Score: tonight, vietnam, taxabl, vietnames, polic, asian, tri
## Topic 6 Top Words:
##       Highest Prob: nation, world, will, must, year, can, peopl
##       FREX: missil, ballist, mutual, scientif, expenditur, freedom, scienc
##       Lift: aggreg, annihil, aria, atlas, ballist, compart, counter
##       Score: compart, ballist, intellectu, steel, possess, imperialist, labor-manag
```

```
# topics 1-6
# difference between different parties
model.stm.ee <- estimateEffect(1:6 ~ party, model.stm, meta = out$meta)
# right associated with cov.value1
plot(model.stm.ee, "party", method="difference", cov.value1="Republican", cov.value2="Democratic")
```




```
# differences between two types
model.stm.ee <- estimateEffect(1:6 ~ sotu_type, model.stm, meta = out$meta)
plot(model.stm.ee, "sotu_type", method="difference", cov.value1="speech", cov.value2="written")
```




```
# The above code is the 20 years data, next we need to find out the first ten years and last ten years
```

```
############################################### first ten year of Vietnam War ##########################
# The following code relates to the first eleven/half year of the Vietnam War

war_first_ten <- metadata[which((metadata$year>=1955) & (metadata$year<=1965)),c(colnames(metadata))]
# Preprocessing of the  data and make it corpus
corpus_sotu_first_ten <- corpus(war_first_ten, text_field = "text")
corpus_sotu_first_ten
```

```
## Corpus consisting of 13 documents and 4 docvars.
## text1 :
## "Mr. President, Mr. Speaker, Members of the Congress:  First,..."
##
## text2 :
## " [Recorded on film and tape and broadcast the same day]  My ..."
##
## text3 :
## " [Read before a joint session by a clerk of the House of Rep..."
##
## text4 :
## "To the Congress of the United States:  I appear before the C..."
##
## text5 :
## "Mr. President, Mr. Speaker, Members of the 85th Congress:  I..."
##
## text6 :
## " [Delivered in person before a joint session]  Mr. President..."
##
## [ reached max_ndoc ... 7 more documents ]
```

```
#Some common pre-processing, remove the punctuation and numbers
toks_first_ten <- tokens(corpus_sotu_first_ten, remove_punct = TRUE, remove_numbers=TRUE)
toks_first_ten <- tokens_wordstem(toks_first_ten)
toks_first_ten <- tokens_select(toks_first_ten,  stopwords("en"), selection = "remove")
dfm_first_ten <- dfm(toks_first_ten)
```

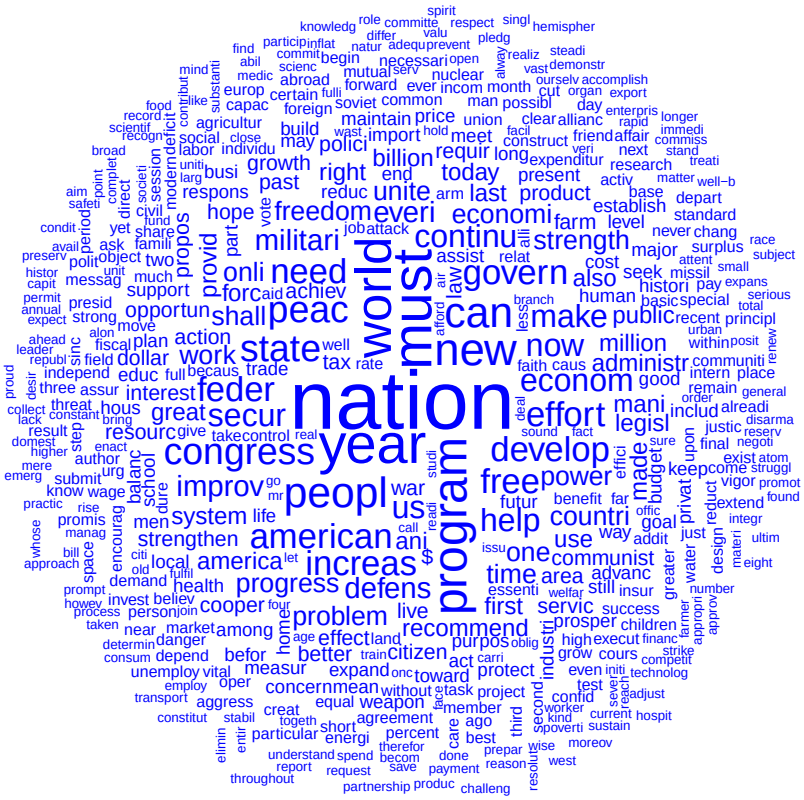
```
dfm_first_ten
```

```
## Document-feature matrix of: 13 documents, 4,155 features (73.42% sparse) and 4 docvars.
##        features
## docs    mr presid speaker member congress first extend cordial greet 84th
##   text1  2      3       1      2       24    13      5       1     1    2
##   text2  0      0       0      0        6     5      0       0     0    0
##   text3  0      0       0      1       29     2      5       0     0    0
##   text4  0      0       0      1       18     7      1       0     0    0
##   text5  2      2       1      3       10     7      2       0     1    0
##   text6  2      3       1      3       16     2      2       0     0    0
## [ reached max_ndoc ... 7 more documents, reached max_nfeat ... 4,145 more features ]
```

```
#Create a document feature matrix (dfm) and trim it with words appeared at least 5%
#toks_first_ten <- corpus_sotu %>%
#  tokens()
#dfm_first_ten <- dfm(toks_first_ten)
dfm_trimmed_first_ten <- dfm_trim(dfm_first_ten, min_docfreq = 0.05, docfreq_type = "prop")
dfm_trimmed_first_ten
```

```
## Document-feature matrix of: 13 documents, 4,155 features (73.42% sparse) and 4 docvars.
##        features
## docs    mr presid speaker member congress first extend cordial greet 84th
##   text1  2      3       1      2       24    13      5       1     1    2
##   text2  0      0       0      0        6     5      0       0     0    0
##   text3  0      0       0      1       29     2      5       0     0    0
##   text4  0      0       0      1       18     7      1       0     0    0
##   text5  2      2       1      3       10     7      2       0     1    0
##   text6  2      3       1      3       16     2      2       0     0    0
## [ reached max_ndoc ... 7 more documents, reached max_nfeat ... 4,145 more features ]
```

```
# word cloud for the first ten years of the Vietnam War
textplot_wordcloud(dfm_trimmed_first_ten, col="blue")
```




```
# Find words that are distinctive between newsletters written by Democrats and Republicans
terms_first_ten_s <- clusterFightinWords(dfm_trimmed_first_ten,war_first_ten$sotu_type == "speech")
sort(terms_first_ten_s, decreasing=T)[1:6] #nation,people
```

```
##  freedom    faith    begin strength     mere      can
## 1.263302 1.215949 1.215949 1.174543 1.157269 1.121395
```

```
terms_first_ten_w <- clusterFightinWords(dfm_trimmed_first_ten,war_first_ten$sotu_type == "written")
sort(terms_first_ten_w, decreasing=T)[1:6]#federal, administration, veterans, education
```

```
## administr   program     eight   veteran      sinc    legisl
##  3.371978  3.158756  2.906599  2.518105  2.495413  2.446863
```

```
############## LDA #############

#Run LDA using quanteda
lda_half <- textmodel_lda(dfm_trimmed_first_ten, k = 6)
#Most likely term for each topic
lda_half.terms <- terms(lda_half, 6)
lda_half.terms
```

```
##      topic1     topic2     topic3      topic4     topic5      topic6
## [1,] "new"      "continu"  "missil"    "independ" "feder"     "nation"
## [2,] "$"        "help"     "soviet"    "month"    "govern"    "must"
## [3,] "million"  "man"      "action"    "domest"   "program"   "world"
## [4,] "tax"      "american" "one"       "urg"      "continu"   "year"
## [5,] "unemploy" "seek"     "real"      "determin" "recommend" "can"
## [6,] "nuclear"  "union"    "technolog" "western"  "legisl"    "peopl"
```

```
#Topical content matrix
mu_half <- lda_half$phi
```

```r
dim(mu_half) #6 topics
```

```
## [1]    6 4155
```

```r
#Most representative words in Topic 1-6
for (i in 1:6){
  print(mu[i,][order(mu_half[i,], decreasing=T)][1:6])
  print("############################################################")
}
```

```
##        itali      coordin        invit     guarante         half        deter
## 8.677166e-06 7.896221e-04 6.160788e-04 1.822205e-04 8.677166e-06 6.160788e-04
## [1] "############################################################"
##        close       expand        littl        three         onli       immens
## 2.922217e-04 3.214439e-03 5.729838e-06 3.157141e-03 5.729838e-06 5.729838e-06
## [1] "############################################################"
##            gear        confront          servic secretary-gener         exhaust
##    7.115665e-06    7.115665e-06    3.422635e-03    7.115665e-06    7.827232e-05
##            eras
##    7.115665e-06
## [1] "############################################################"
##    establish          men       commit         alli       allevi      germani
## 1.359897e-05 5.045217e-03 1.359897e-05 1.359897e-05 1.359897e-05 2.855783e-04
## [1] "############################################################"
##       justic      freedom      mention        close         back        reduc
## 1.375209e-03 1.238927e-05 1.238927e-05 1.238927e-05 1.238927e-05 3.853063e-03
## [1] "############################################################"
##       common     interest         form         just      factori        shape
## 4.746422e-06 1.471391e-04 4.746422e-06 3.801884e-03 4.746422e-06 1.001495e-03
## [1] "############################################################"
```

```r
#Topical prevalence matrix
pi_half <- lda_half$theta
dim(pi_half)
```

```
## [1] 13  6
```

```r
#Most representative documents in Topic 1-10
for (i in 1:6){
  print(war_first_ten[order(pi_half[i,],decreasing=T),])
  print("############################################################")
}
```

```
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 2 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## 3 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 4 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## 5 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 6 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## [1] "############################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
```

```
## 1 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 2 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## 3 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 4 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## 5 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 6 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 2 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## 3 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 4 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 5 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## 6 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 2 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## 3 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 4 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 5 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 6 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 2 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 3 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## 4 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 5 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 6 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## [1] "##########################################################"
## # A tibble: 6 x 5
##   president             year party      sotu_type text
##   <chr>                <dbl> <chr>      <chr>     <chr>
## 1 Dwight D. Eisenhower  1959 Republican speech    "\n\n[Delivered in person bef~
## 2 Dwight D. Eisenhower  1958 Republican speech    "Mr. President, Mr. Speaker, ~
## 3 Dwight D. Eisenhower  1956 Republican written   "\n\n[Read before a joint ses~
## 4 Dwight D. Eisenhower  1956 Republican speech    "\n\n[Recorded on film and ta~
## 5 Dwight D. Eisenhower  1957 Republican speech    "To the Congress of the Unite~
## 6 Dwight D. Eisenhower  1955 Republican speech    "Mr. President, Mr. Speaker, ~
## [1] "##########################################################"
```

```
############ STM ############

# Structural Topic Model
#STM
#library(tm)
temp_first_ten <-textProcessor(documents=war_first_ten$text,metadata=war_first_ten)
```

```
## Building corpus...
## Converting to Lower Case...
## Removing punctuation...
## Removing stopwords...
## Removing numbers...
## Stemming...
## Creating Output...
```

```
out <- prepDocuments(temp_first_ten$documents, temp_first_ten$vocab, temp_first_ten$meta)
```

```
## Removing 2110 of 4415 terms (2110 of 14305 tokens) due to frequency
## Your corpus now has 13 documents, 2305 terms and 12195 tokens.
```

```
# Run the model
# Should we change the value of K
model.stm <- stm(out$documents, out$vocab, K = 6, prevalence = ~ party,
                 data = out$meta, max.em.its = 10)
```

```
## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      ......
##   Recovering initialization...
##      ......................
## Initialization complete.
## ............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -6.781)
## ............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 2 (approx. per word bound = -6.726, relative change = 8.115e-03)
## ............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 3 (approx. per word bound = -6.708, relative change = 2.654e-03)
## ............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 4 (approx. per word bound = -6.700, relative change = 1.259e-03)
## ............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 5 (approx. per word bound = -6.694, relative change = 7.747e-04)
## Topic 1: will, program, year, nation, congress
##  Topic 2: must, will, nation, new, can
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, must, peopl, year
##  Topic 6: program, nation, year, feder, new
## ............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 6 (approx. per word bound = -6.691, relative change = 5.500e-04)
```

```
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 7 (approx. per word bound = -6.688, relative change = 4.587e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 8 (approx. per word bound = -6.685, relative change = 3.910e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 9 (approx. per word bound = -6.683, relative change = 3.033e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Model Terminated Before Convergence Reached
```

```
model.stm <- stm(out$documents, out$vocab, K = 6, prevalence = ~ party,
                 data = out$meta)
```

```
## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      ......
##   Recovering initialization...
##      ......................
## Initialization complete.
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -6.781)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 2 (approx. per word bound = -6.726, relative change = 8.115e-03)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 3 (approx. per word bound = -6.708, relative change = 2.654e-03)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 4 (approx. per word bound = -6.700, relative change = 1.259e-03)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 5 (approx. per word bound = -6.694, relative change = 7.747e-04)
## Topic 1: will, program, year, nation, congress
##  Topic 2: must, will, nation, new, can
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, must, peopl, year
##  Topic 6: program, nation, year, feder, new
## .............
## Completed E-Step (0 seconds).
```

```
## Completed M-Step.
## Completing Iteration 6 (approx. per word bound = -6.691, relative change = 5.500e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 7 (approx. per word bound = -6.688, relative change = 4.587e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 8 (approx. per word bound = -6.685, relative change = 3.910e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 9 (approx. per word bound = -6.683, relative change = 3.033e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 10 (approx. per word bound = -6.681, relative change = 2.516e-04)
## Topic 1: will, program, year, nation, congress
##  Topic 2: will, must, nation, new, year
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, must, peopl, can
##  Topic 6: program, nation, feder, year, new
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 11 (approx. per word bound = -6.680, relative change = 2.180e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 12 (approx. per word bound = -6.679, relative change = 1.952e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 13 (approx. per word bound = -6.677, relative change = 1.755e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 14 (approx. per word bound = -6.676, relative change = 1.632e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 15 (approx. per word bound = -6.675, relative change = 1.471e-04)
## Topic 1: will, program, year, nation, must
##  Topic 2: will, nation, must, new, world
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, must, can, world
##  Topic 6: program, nation, year, feder, new
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 16 (approx. per word bound = -6.674, relative change = 1.234e-04)
```

```
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 17 (approx. per word bound = -6.674, relative change = 1.085e-04)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 18 (approx. per word bound = -6.673, relative change = 9.894e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 19 (approx. per word bound = -6.672, relative change = 9.020e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 20 (approx. per word bound = -6.672, relative change = 8.380e-05)
## Topic 1: will, program, year, nation, must
##  Topic 2: will, nation, must, new, world
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, must, world, can
##  Topic 6: program, nation, year, feder, new
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 21 (approx. per word bound = -6.671, relative change = 7.866e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 22 (approx. per word bound = -6.671, relative change = 7.513e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 23 (approx. per word bound = -6.670, relative change = 7.017e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 24 (approx. per word bound = -6.670, relative change = 6.384e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 25 (approx. per word bound = -6.670, relative change = 5.750e-05)
## Topic 1: will, program, year, nation, must
##  Topic 2: will, nation, must, new, world
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, must, world, can
##  Topic 6: program, nation, year, feder, new
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 26 (approx. per word bound = -6.669, relative change = 5.274e-05)
## .............
## Completed E-Step (0 seconds).
```

```
## Completed M-Step.
## Completing Iteration 27 (approx. per word bound = -6.669, relative change = 5.028e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 28 (approx. per word bound = -6.669, relative change = 4.935e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 29 (approx. per word bound = -6.668, relative change = 4.917e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 30 (approx. per word bound = -6.668, relative change = 5.033e-05)
## Topic 1: will, program, year, nation, must
##  Topic 2: will, nation, must, new, world
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, must, world, can
##  Topic 6: nation, program, year, feder, new
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 31 (approx. per word bound = -6.668, relative change = 4.915e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 32 (approx. per word bound = -6.667, relative change = 3.924e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 33 (approx. per word bound = -6.667, relative change = 3.384e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 34 (approx. per word bound = -6.667, relative change = 3.249e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 35 (approx. per word bound = -6.667, relative change = 3.198e-05)
## Topic 1: will, program, year, nation, must
##  Topic 2: will, nation, must, new, year
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, must, world, can
##  Topic 6: nation, program, year, feder, new
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 36 (approx. per word bound = -6.666, relative change = 3.302e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 37 (approx. per word bound = -6.666, relative change = 3.408e-05)
```

```
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 38 (approx. per word bound = -6.666, relative change = 3.544e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 39 (approx. per word bound = -6.666, relative change = 3.615e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 40 (approx. per word bound = -6.665, relative change = 3.528e-05)
## Topic 1: will, program, year, nation, must
##  Topic 2: will, nation, must, new, year
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, world, must, can
##  Topic 6: nation, program, year, feder, new
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 41 (approx. per word bound = -6.665, relative change = 3.096e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 42 (approx. per word bound = -6.665, relative change = 2.785e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 43 (approx. per word bound = -6.665, relative change = 2.561e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 44 (approx. per word bound = -6.665, relative change = 2.402e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 45 (approx. per word bound = -6.665, relative change = 2.268e-05)
## Topic 1: will, program, nation, year, must
##  Topic 2: will, nation, must, new, year
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, world, must, can
##  Topic 6: nation, program, year, new, feder
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 46 (approx. per word bound = -6.664, relative change = 2.146e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 47 (approx. per word bound = -6.664, relative change = 2.266e-05)
## .............
## Completed E-Step (0 seconds).
```

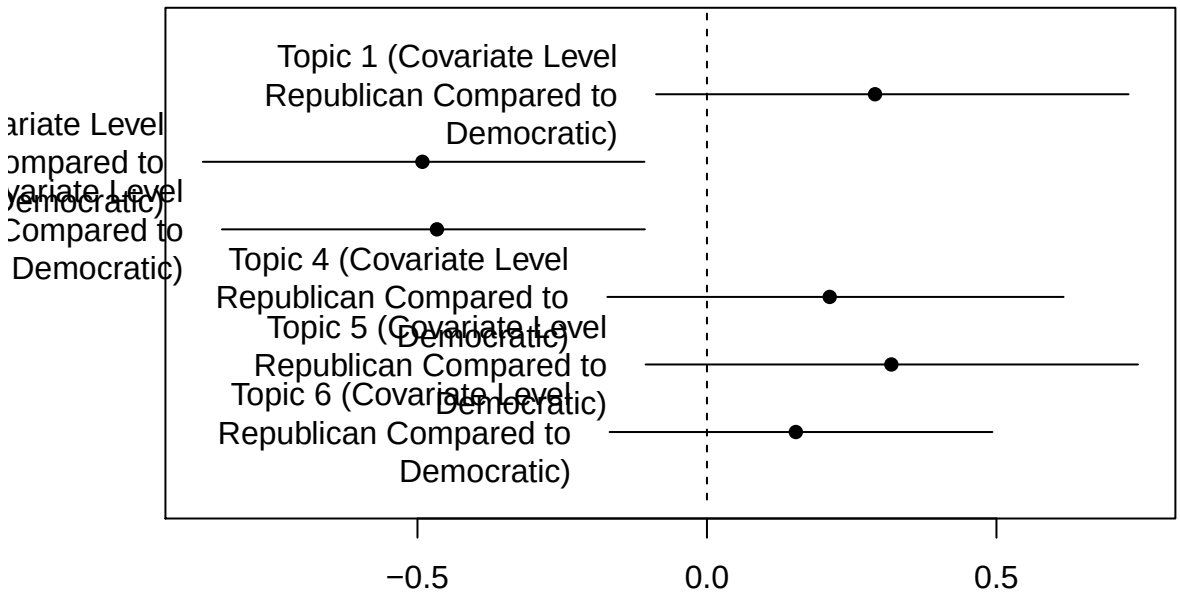
```
## Completed M-Step.
## Completing Iteration 48 (approx. per word bound = -6.664, relative change = 2.143e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 49 (approx. per word bound = -6.664, relative change = 1.921e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 50 (approx. per word bound = -6.664, relative change = 1.908e-05)
## Topic 1: will, program, nation, year, must
##  Topic 2: will, nation, must, new, year
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, world, must, can
##  Topic 6: nation, program, year, new, feder
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 51 (approx. per word bound = -6.664, relative change = 2.107e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 52 (approx. per word bound = -6.664, relative change = 2.644e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 53 (approx. per word bound = -6.663, relative change = 2.247e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 54 (approx. per word bound = -6.663, relative change = 2.378e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 55 (approx. per word bound = -6.663, relative change = 3.799e-05)
## Topic 1: will, nation, program, year, must
##  Topic 2: will, nation, must, new, year
##  Topic 3: will, nation, year, new, world
##  Topic 4: nation, world, will, must, peac
##  Topic 5: nation, will, world, must, can
##  Topic 6: nation, program, year, new, feder
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 56 (approx. per word bound = -6.663, relative change = 3.719e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 57 (approx. per word bound = -6.663, relative change = 2.547e-05)
## .............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 58 (approx. per word bound = -6.663, relative change = 1.710e-05)
```

```
## ............
## Completed E-Step (0 seconds).
## Completed M-Step.
## Model Converged
```

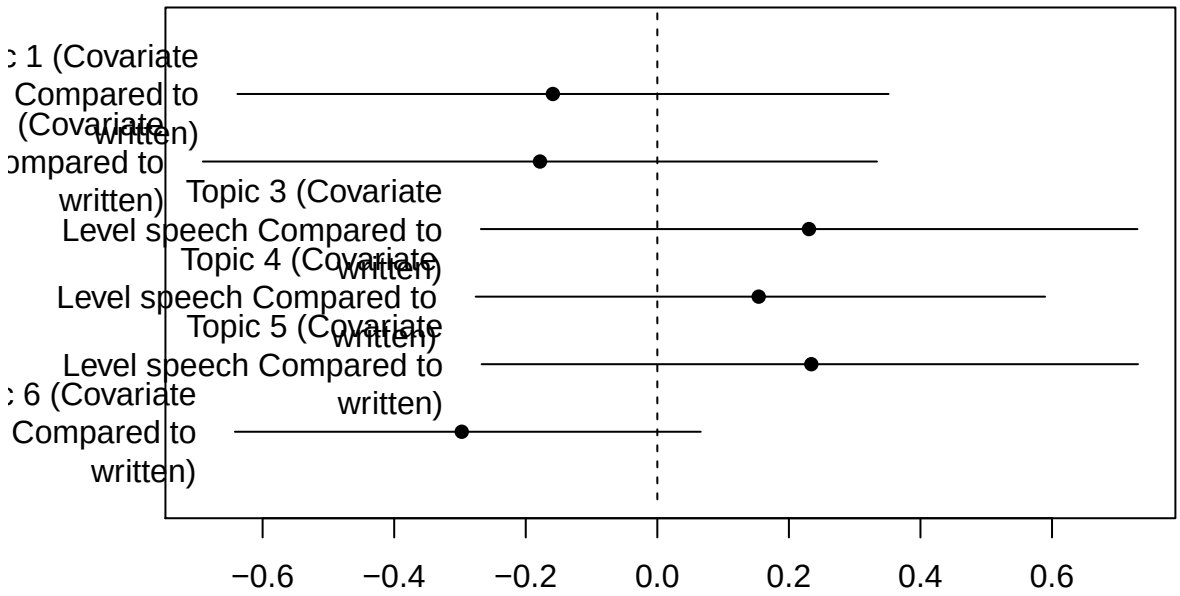
```
labelTopics(model.stm)
```

```
## Topic 1 Top Words:
##       Highest Prob: will, nation, program, year, must, congress, feder
##       FREX: recommend, legisl, measur, shall, hous, highway, exist
##       Lift: acreag, equit, pictur, plain, self, accompani, wartim
##       Score: pictur, legisl, mention, district, survivor, equit, disast
## Topic 2 Top Words:
##       Highest Prob: will, nation, must, new, year, world, can
##       FREX: job, bill, lack, session, unemploy, instead, member
##       Lift: actual, draw, escap, hate, hungri, sister, someday
##       Score: hate, lack, bill, gold, instead, tool, unemploy
## Topic 3 Top Words:
##       Highest Prob: will, nation, year, new, world, american, can
##       FREX: begin, allianc, europ, societi, choic, centuri, citi
##       Lift: coloni, elector, empir, hill, recreat, street, turbul
##       Score: hill, choic, viet-nam, choos, run, recreat, elector
## Topic 4 Top Words:
##       Highest Prob: nation, world, will, must, peac, peopl, year
##       FREX: missil, soviet, necessari, kind, one, becom, real
##       Lift: encroach, featur, intellig, newly-develop, tempt, annihil, cement
##       Score: shield, convict, intellig, ballist, jurisdict, scientist, featur
## Topic 5 Top Words:
##       Highest Prob: nation, will, world, must, can, year, peopl
##       FREX: price, expenditur, human, field, law, respons, recent
##       Lift: defin, entiti, uphold, amaz, ceaseless, foreshadow, durabl
##       Score: entiti, legisl, uphold, defin, armament, self-disciplin, exceed
## Topic 6 Top Words:
##       Highest Prob: nation, program, year, new, feder, govern, develop
##       FREX: eight, veteran, sinc, percent, vigor, strengthen, advanc
##       Lift: aviat, largest, reclam, australia, cornerston, egypt, government-own
##       Score: largest, veteran, eight, survivor, factor, readjust, becam
```

```
# topics 1-10
model.stm.ee <- estimateEffect(1:6 ~ party, model.stm, meta = out$meta)
# right associated with cov.value1
plot(model.stm.ee, "party", method="difference", cov.value1="Republican", cov.value2="Democratic")
```




```
model.stm.ee <- estimateEffect(1:6 ~ sotu_type, model.stm, meta = out$meta)
plot(model.stm.ee, "sotu_type", method="difference", cov.value1="speech", cov.value2="written")
```




```
############################################## last ten year of Vietnam War ##########################
# The following code relates to the last ten year of the Vietnam War
war_last_ten <- metadata[which((metadata$year>=1966) & (metadata$year<=1975)),c(colnames(metadata))]
# Preprocessing of the  data and make it corpus
corpus_sotu_last_ten <- corpus(war_last_ten, text_field = "text")
corpus_sotu_last_ten
```

```
## Corpus consisting of 11 documents and 4 docvars.
## text1 :
## " [ Delivered in person before a joint session at 9:04 p.m. ]..."
##
## text2 :
## " [Delivered in person before a joint session at 9:33 p.m.]  ..."
##
```

```
## text3 :
## " [Delivered in person before a joint session at 9:05 p.m.]  ..."
##
## text4 :
## "Mr. Speaker, Mr. President, Members of the Congress and my f..."
##
## text5 :
## " Mr. Speaker, Mr. President, my colleagues in the Congress, ..."
##
## text6 :
## " Mr. Speaker, Mr. President, my colleagues in the Congress, ..."
##
## [ reached max_ndoc ... 5 more documents ]
```

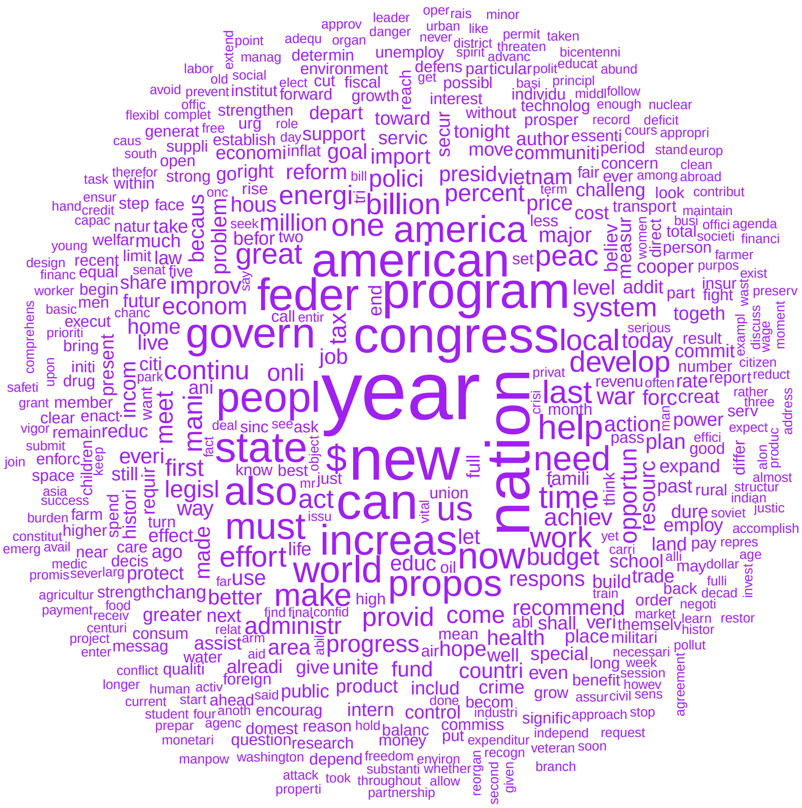
```
#Some common pre-processing, remove the punctuation and numbers
toks_last_ten <- tokens(corpus_sotu_last_ten, remove_punct = TRUE, remove_numbers=TRUE)
toks_last_ten <- tokens_wordstem(toks_last_ten)
toks_last_ten <- tokens_select(toks_last_ten,  stopwords("en"), selection = "remove")
dfm <- dfm(toks_last_ten)
dfm
```

```
## Document-feature matrix of: 11 documents, 4,086 features (71.76% sparse) and 4 docvars.
##        features
## docs    deliv person befor joint session p.m mr speaker presid member
##   text1     1      2     8     1       1   1  2       1      9      6
##   text2     1      3     9     1       2   1  3       1     14      5
##   text3     1      3     8     1       3   1  3       1      6      1
##   text4     0      5     4     0       0   0  4       3     15      5
##   text5     0      2     3     1       2   0  2       1      4      2
##   text6     0      2     4     0       2   0  4       5      3      4
## [ reached max_ndoc ... 5 more documents, reached max_nfeat ... 4,076 more features ]
```

```
#Create a document feature matrix (dfm) and trim it with words appeared at least 5%
#toks_last_ten <- corpus_sotu %>%
#  tokens()
dfm_last_ten <- dfm(toks_last_ten)
dfm_trimmed_last_ten <- dfm_trim(dfm_last_ten, min_docfreq = 0.05, docfreq_type = "prop")
dfm_trimmed_last_ten
```

```
## Document-feature matrix of: 11 documents, 4,086 features (71.76% sparse) and 4 docvars.
##        features
## docs    deliv person befor joint session p.m mr speaker presid member
##   text1     1      2     8     1       1   1  2       1      9      6
##   text2     1      3     9     1       2   1  3       1     14      5
##   text3     1      3     8     1       3   1  3       1      6      1
##   text4     0      5     4     0       0   0  4       3     15      5
##   text5     0      2     3     1       2   0  2       1      4      2
##   text6     0      2     4     0       2   0  4       5      3      4
## [ reached max_ndoc ... 5 more documents, reached max_nfeat ... 4,076 more features ]
```

```
# word cloud for the first ten years of the Vietnam War
textplot_wordcloud(dfm_trimmed_last_ten, col="purple")
```




```r
# Find words that are distinctive between newsletters written by Democrats and Republicans
terms_last_ten_s <- clusterFightinWords(dfm_trimmed_last_ten,war_last_ten$sotu_type == "speech")
sort(terms_last_ten_s, decreasing=T)[1:6] #nation,people
```

```
##    peopl  america      let       us    great  tonight
## 4.889528 4.341263 4.102573 3.686946 3.666421 3.528308
```

```r
terms_last_ten_s <- clusterFightinWords(dfm_trimmed_last_ten,war_last_ten$sotu_type == "written")
sort(terms_last_ten_s, decreasing=T)[1:6]#federal, administration, veterans, education
```

```
##       also administr     feder      educ    servic      dure
##   3.672145  2.760344  2.621418  2.570444  2.418945  2.347972
```

```r
############# LDA ############

#Run LDA using quanteda
lda <- textmodel_lda(dfm_trimmed_last_ten, k = 6)
#Most likely term for each topic
lda.terms <- terms(lda, 6)
lda.terms
```

```
##      topic1     topic2    topic3    topic4    topic5     topic6
## [1,] "year"     "peopl"   "energi"  "feder"   "program"  "vietnam"
## [2,] "nation"   "america" "oil"     "also"    "new"      "tonight"
## [3,] "must"     "us"      "tax"     "develop" "also"     "hope"
## [4,] "$"        "new"     "suppli"  "need"    "share"    "think"
## [5,] "congress" "govern"  "product" "propos"  "employ"   "tri"
## [6,] "help"     "great"   "cut"     "govern"  "research" "men"
```

```r
#Topical content matrix
mu_last_ten <- lda$phi
dim(mu_last_ten)
```

```
## [1]    6 4086
```

```r
mu_last_ten[1:6,1:20]
```

```
##               deliv       person        befor        joint      session
## topic1 6.397707e-06 2.053664e-03 3.397182e-03 6.461684e-04 1.861733e-03
## topic2 8.411440e-04 8.411440e-04 1.765478e-03 9.243340e-06 1.941101e-04
## topic3 4.867127e-05 4.867127e-05 4.867127e-05 4.867127e-05 4.867127e-05
## topic4 9.436099e-06 1.981581e-04 9.436099e-06 9.436099e-06 9.436099e-06
## topic5 2.529340e-05 2.048766e-03 2.529340e-05 2.529340e-05 2.529340e-05
## topic6 3.574214e-04 3.249285e-05 6.823499e-04 3.574214e-04 3.249285e-05
##                 p.m           mr      speaker       presid       member
## topic1 6.397707e-06 1.030031e-03 6.397707e-06 6.397707e-06 7.741226e-04
## topic2 9.243340e-06 1.026011e-03 1.765478e-03 6.017414e-03 3.244412e-03
## topic3 4.867127e-05 4.867127e-05 4.867127e-05 5.353840e-04 4.867127e-05
## topic4 9.436099e-06 9.436099e-06 9.436099e-06 9.436099e-06 9.436099e-06
## topic5 2.782274e-04 2.529340e-05 2.529340e-05 2.529340e-05 2.529340e-05
## topic6 6.823499e-04 3.249285e-05 3.249285e-05 9.780348e-03 3.249285e-05
##                hous        senat       fellow     american         come
## topic1 2.949343e-03 5.182143e-04 6.397707e-06 1.254590e-02 1.925710e-03
## topic2 9.243340e-06 2.135212e-03 3.789769e-04 9.807184e-03 7.034182e-03
## topic3 4.867127e-05 4.867127e-05 4.867127e-05 4.867127e-05 4.867127e-05
## topic4 3.972598e-03 3.868800e-04 9.436099e-06 9.436099e-06 9.436099e-06
## topic5 2.529340e-05 2.529340e-05 2.529340e-05 2.529340e-05 2.529340e-05
## topic6 3.249285e-05 3.249285e-05 2.631921e-03 3.574214e-04 3.249285e-05
##             tonight       report        state        union        third
## topic1 7.037478e-05 2.437526e-03 3.972976e-03 2.053664e-03 6.397707e-06
## topic2 9.335774e-04 2.865435e-04 7.496349e-03 2.320078e-03 1.765478e-03
## topic3 4.867127e-05 4.867127e-05 4.867127e-05 4.867127e-05 4.867127e-05
## topic4 9.436099e-06 9.436099e-06 1.038914e-02 9.436099e-06 9.436099e-06
## topic5 2.529340e-05 2.529340e-05 2.529340e-05 2.529340e-05 2.529340e-05
## topic6 2.180270e-02 3.249285e-05 1.657135e-03 3.249285e-05 3.249285e-05
```

```r
#Most representative words in Topic 1-10
for (i in 1:6){
  print(mu_last_ten[i,][order(mu_last_ten[i,], decreasing=T)][1:6])
  print("##########################################################")
}
```

```
##       year     nation       must          $   congress       help
## 0.02495746 0.01907156 0.01427328 0.01356954 0.01273783 0.01273783
## [1] "##########################################################"
##      peopl    america         us        new     govern      great
## 0.02357976 0.01988242 0.01905052 0.01516832 0.01489102 0.01470615
## [1] "##########################################################"
##     energi        oil        tax     suppli    product        cut
## 0.02973815 0.02146403 0.01562348 0.01513677 0.01221649 0.01026964
## [1] "##########################################################"
##      feder       also    develop       need     propos     govern
## 0.02161810 0.01321997 0.01227636 0.01133275 0.01076659 0.01057787
## [1] "##########################################################"
```

```
##     program         new        also       share      employ    research
## 0.024306961 0.023042291 0.019754148 0.011407325 0.009130919 0.008625051
## [1] "######################################################"
##    vietnam   tonight      hope     think        tri        men
## 0.02212763 0.02180270 0.01335456 0.01270470 0.01140499 0.01075513
## [1] "######################################################"
```

```
#Topical prevalence matrix
pi <- lda$theta
#Most representative documents in Topic 1-10
for (i in 1:6){
  print(war_last_ten[order(pi[i,],decreasing=T),])
  print("######################################################")
}
```

```
## # A tibble: 6 x 5
##   president          year party      sotu_type text
##   <chr>             <dbl> <chr>      <chr>     <chr>
## 1 Lyndon B. Johnson  1966 Democratic speech    "\n\n[ Delivered in person befor~
## 2 Richard M. Nixon   1971 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 3 Lyndon B. Johnson  1967 Democratic speech    "\n\n[Delivered in person before~
## 4 Lyndon B. Johnson  1969 Democratic written   "Mr. Speaker, Mr. President, Mem~
## 5 Richard M. Nixon   1970 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 6 Lyndon B. Johnson  1968 Democratic speech    "\n\n[Delivered in person before~
## [1] "######################################################"
## # A tibble: 6 x 5
##   president          year party      sotu_type text
##   <chr>             <dbl> <chr>      <chr>     <chr>
## 1 Lyndon B. Johnson  1966 Democratic speech    "\n\n[ Delivered in person befor~
## 2 Lyndon B. Johnson  1967 Democratic speech    "\n\n[Delivered in person before~
## 3 Richard M. Nixon   1971 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 4 Lyndon B. Johnson  1969 Democratic written   "Mr. Speaker, Mr. President, Mem~
## 5 Richard M. Nixon   1970 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 6 Lyndon B. Johnson  1968 Democratic speech    "\n\n[Delivered in person before~
## [1] "######################################################"
## # A tibble: 6 x 5
##   president          year party      sotu_type text
##   <chr>             <dbl> <chr>      <chr>     <chr>
## 1 Lyndon B. Johnson  1966 Democratic speech    "\n\n[ Delivered in person befor~
## 2 Lyndon B. Johnson  1967 Democratic speech    "\n\n[Delivered in person before~
## 3 Richard M. Nixon   1971 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 4 Lyndon B. Johnson  1969 Democratic written   "Mr. Speaker, Mr. President, Mem~
## 5 Richard M. Nixon   1970 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 6 Lyndon B. Johnson  1968 Democratic speech    "\n\n[Delivered in person before~
## [1] "######################################################"
## # A tibble: 6 x 5
##   president          year party      sotu_type text
##   <chr>             <dbl> <chr>      <chr>     <chr>
## 1 Lyndon B. Johnson  1966 Democratic speech    "\n\n[ Delivered in person befor~
## 2 Lyndon B. Johnson  1967 Democratic speech    "\n\n[Delivered in person before~
## 3 Richard M. Nixon   1971 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 4 Lyndon B. Johnson  1969 Democratic written   "Mr. Speaker, Mr. President, Mem~
## 5 Richard M. Nixon   1970 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 6 Lyndon B. Johnson  1968 Democratic speech    "\n\n[Delivered in person before~
## [1] "######################################################"
```

```
## # A tibble: 6 x 5
##   president          year party      sotu_type text
##   <chr>             <dbl> <chr>      <chr>     <chr>
## 1 Lyndon B. Johnson  1967 Democratic speech    "\n\n[Delivered in person before~
## 2 Lyndon B. Johnson  1966 Democratic speech    "\n\n[ Delivered in person befor~
## 3 Lyndon B. Johnson  1969 Democratic written   "Mr. Speaker, Mr. President, Mem~
## 4 Richard M. Nixon   1970 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 5 Richard M. Nixon   1971 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 6 Lyndon B. Johnson  1968 Democratic speech    "\n\n[Delivered in person before~
## [1] "############################################################"
## # A tibble: 6 x 5
##   president          year party      sotu_type text
##   <chr>             <dbl> <chr>      <chr>     <chr>
## 1 Lyndon B. Johnson  1967 Democratic speech    "\n\n[Delivered in person before~
## 2 Lyndon B. Johnson  1966 Democratic speech    "\n\n[ Delivered in person befor~
## 3 Lyndon B. Johnson  1969 Democratic written   "Mr. Speaker, Mr. President, Mem~
## 4 Richard M. Nixon   1970 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 5 Richard M. Nixon   1971 Republican speech    "\n\nMr. Speaker, Mr. President,~
## 6 Lyndon B. Johnson  1968 Democratic speech    "\n\n[Delivered in person before~
## [1] "############################################################"
```

```
############ STM ############
# Structural Topic Model
#STM
#library(tm)
temp_last_ten <-textProcessor(documents=war_last_ten$text,metadata=war_last_ten)
```

```
## Building corpus...
## Converting to Lower Case...
## Removing punctuation...
## Removing stopwords...
## Removing numbers...
## Stemming...
## Creating Output...
```

```
out <- prepDocuments(temp_last_ten$documents, temp_last_ten$vocab, temp_last_ten$meta)
```

```
## Removing 2244 of 4434 terms (2244 of 12768 tokens) due to frequency
## Your corpus now has 11 documents, 2190 terms and 10524 tokens.
```

```
# Run the model
# Should we change the value of K
model.stm <- stm(out$documents, out$vocab, K = 6, prevalence = ~ party,
                 data = out$meta, max.em.its = 10)
```

```
## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      ......
##   Recovering initialization...
##      ......................
## Initialization complete.
## ...........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -6.648)
```

```
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 2 (approx. per word bound = -6.594, relative change = 8.092e-03)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 3 (approx. per word bound = -6.567, relative change = 4.186e-03)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 4 (approx. per word bound = -6.554, relative change = 1.977e-03)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 5 (approx. per word bound = -6.548, relative change = 8.561e-04)
## Topic 1: govern, will, peopl, can, congress
##  Topic 2: will, feder, year, program, new
##  Topic 3: year, will, congress, nation, can
##  Topic 4: will, year, america, nation, american
##  Topic 5: will, nation, year, peopl, must
##  Topic 6: will, new, year, program, can
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 6 (approx. per word bound = -6.545, relative change = 4.047e-04)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 7 (approx. per word bound = -6.544, relative change = 2.170e-04)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 8 (approx. per word bound = -6.543, relative change = 1.196e-04)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 9 (approx. per word bound = -6.543, relative change = 8.005e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Model Terminated Before Convergence Reached
```

```
model.stm <- stm(out$documents, out$vocab, K = 6, prevalence = ~ party,
                 data = out$meta)
```

```
## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##      ......
##   Recovering initialization...
##      ..................
## Initialization complete.
## ..........
## Completed E-Step (0 seconds).
```

```
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -6.648)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 2 (approx. per word bound = -6.594, relative change = 8.092e-03)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 3 (approx. per word bound = -6.567, relative change = 4.186e-03)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 4 (approx. per word bound = -6.554, relative change = 1.977e-03)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 5 (approx. per word bound = -6.548, relative change = 8.561e-04)
## Topic 1: govern, will, peopl, can, congress
##  Topic 2: will, feder, year, program, new
##  Topic 3: year, will, congress, nation, can
##  Topic 4: will, year, america, nation, american
##  Topic 5: will, nation, year, peopl, must
##  Topic 6: will, new, year, program, can
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 6 (approx. per word bound = -6.545, relative change = 4.047e-04)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 7 (approx. per word bound = -6.544, relative change = 2.170e-04)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 8 (approx. per word bound = -6.543, relative change = 1.196e-04)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 9 (approx. per word bound = -6.543, relative change = 8.005e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 10 (approx. per word bound = -6.542, relative change = 6.917e-05)
## Topic 1: govern, will, peopl, can, congress
##  Topic 2: will, feder, year, new, program
##  Topic 3: year, will, congress, nation, can
##  Topic 4: will, year, america, nation, american
##  Topic 5: will, nation, year, peopl, must
##  Topic 6: will, new, year, program, can
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 11 (approx. per word bound = -6.542, relative change = 5.964e-05)
```

```
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 12 (approx. per word bound = -6.542, relative change = 5.287e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 13 (approx. per word bound = -6.541, relative change = 4.561e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 14 (approx. per word bound = -6.541, relative change = 3.873e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 15 (approx. per word bound = -6.541, relative change = 3.320e-05)
## Topic 1: govern, will, peopl, can, congress
##  Topic 2: will, feder, year, new, program
##  Topic 3: year, will, congress, nation, can
##  Topic 4: will, year, america, nation, american
##  Topic 5: will, nation, year, peopl, must
##  Topic 6: will, new, year, program, also
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 16 (approx. per word bound = -6.541, relative change = 2.980e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 17 (approx. per word bound = -6.540, relative change = 2.937e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 18 (approx. per word bound = -6.540, relative change = 3.063e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 19 (approx. per word bound = -6.540, relative change = 3.263e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 20 (approx. per word bound = -6.540, relative change = 3.342e-05)
## Topic 1: govern, will, peopl, can, congress
##  Topic 2: will, feder, year, new, program
##  Topic 3: year, will, congress, nation, can
##  Topic 4: will, year, america, nation, american
##  Topic 5: will, nation, year, peopl, must
##  Topic 6: will, new, year, program, also
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 21 (approx. per word bound = -6.540, relative change = 3.365e-05)
## ..........
## Completed E-Step (0 seconds).
```

```
## Completed M-Step.
## Completing Iteration 22 (approx. per word bound = -6.539, relative change = 3.327e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 23 (approx. per word bound = -6.539, relative change = 3.249e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 24 (approx. per word bound = -6.539, relative change = 3.108e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 25 (approx. per word bound = -6.539, relative change = 2.797e-05)
## Topic 1: govern, will, peopl, can, congress
##  Topic 2: will, feder, year, new, program
##  Topic 3: year, will, congress, nation, can
##  Topic 4: will, year, america, nation, american
##  Topic 5: will, nation, year, peopl, must
##  Topic 6: will, new, year, program, also
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 26 (approx. per word bound = -6.539, relative change = 2.458e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 27 (approx. per word bound = -6.538, relative change = 2.265e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 28 (approx. per word bound = -6.538, relative change = 2.071e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 29 (approx. per word bound = -6.538, relative change = 1.952e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 30 (approx. per word bound = -6.538, relative change = 2.088e-05)
## Topic 1: govern, will, peopl, can, congress
##  Topic 2: will, feder, year, new, program
##  Topic 3: year, will, congress, nation, can
##  Topic 4: will, year, america, nation, new
##  Topic 5: will, nation, year, peopl, must
##  Topic 6: will, new, year, program, also
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 31 (approx. per word bound = -6.538, relative change = 2.176e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 32 (approx. per word bound = -6.538, relative change = 2.387e-05)
```

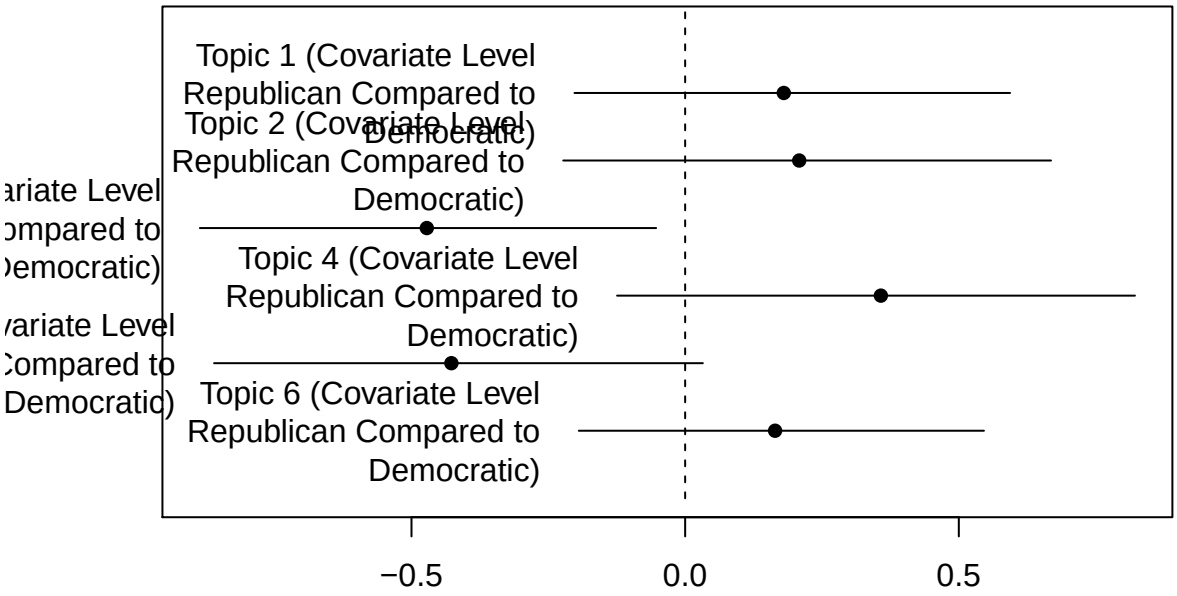
```
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 33 (approx. per word bound = -6.538, relative change = 2.265e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 34 (approx. per word bound = -6.537, relative change = 1.854e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 35 (approx. per word bound = -6.537, relative change = 1.735e-05)
## Topic 1: govern, will, peopl, can, congress
##  Topic 2: will, feder, year, new, program
##  Topic 3: year, will, congress, nation, can
##  Topic 4: will, year, america, nation, new
##  Topic 5: will, nation, year, peopl, must
##  Topic 6: will, new, year, program, also
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 36 (approx. per word bound = -6.537, relative change = 3.056e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 37 (approx. per word bound = -6.537, relative change = 4.334e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 38 (approx. per word bound = -6.537, relative change = 3.837e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Completing Iteration 39 (approx. per word bound = -6.536, relative change = 2.680e-05)
## ..........
## Completed E-Step (0 seconds).
## Completed M-Step.
## Model Converged
```

```
labelTopics(model.stm)
```

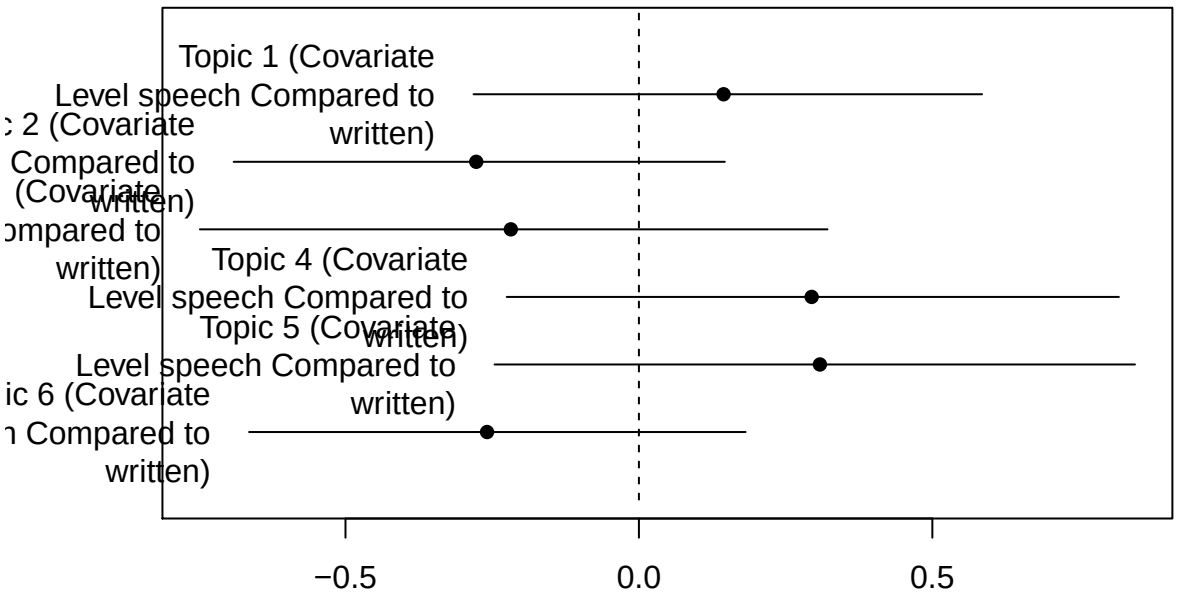
```
## Topic 1 Top Words:
##       Highest Prob: govern, will, peopl, can, congress, new, nation
##       FREX: open, chanc, door, govern, function, vision, washington
##       Lift: rob, abolish, function, patron, privileg, vision, bigger
##       Score: function, vision, truli, privileg, seventi, rob, escap
## Topic 2 Top Words:
##       Highest Prob: will, feder, year, new, program, increas, nation
##       FREX: oil, energi, suppli, environment, ensur, legisl, activ
##       Lift: automat, automot, borrow, consumpt, decontrol, disloc, elig
##       Score: oil, petroleum, environment, bicentenni, suppli, ensur, veteran
## Topic 3 Top Words:
##       Highest Prob: year, will, congress, nation, can, american, billion
##       FREX: think, believ, pass, hope, treati, billion, talk
##       Lift: ban, chairman, clearer, hall, nixon, told, gold
```

```
##       Score: nixon, tonight, polic, salari, know, ban, chairman
## Topic 4 Top Words:
##       Highest Prob: will, year, america, nation, new, american, world
##       FREX: look, peac, toward, america, differ, one, messag
##       Lift: investig, refer, worldth, deeper, partisanship, perfect, superior
##       Score: investig, recal, truli, look, chamber, colleagu, tonight
## Topic 5 Top Words:
##       Highest Prob: will, nation, year, peopl, must, can, american
##       FREX: vietnam, south, tri, alon, attack, men, tonight
##       Lift: -tax, ambit, contin, outlaw, preliminari, slum, conquest
##       Score: servant, tonight, north, wish, aggress, conquest, south
## Topic 6 Top Words:
##       Highest Prob: will, new, year, program, also, can, increas
##       FREX: expand, field, also, older, research, substanti, influenc
##       Lift: satisfi, spiritu, arriv, dealt, discoveri, sell, suburb
##       Score: spiritu, research, career, expand, reform, environment, older
```

```
# topics 1-10
model.stm.ee <- estimateEffect(1:6 ~ party, model.stm, meta = out$meta)
# right associated with cov.value1
plot(model.stm.ee, "party", method="difference", cov.value1="Republican", cov.value2="Democratic")
```




```
model.stm.ee <- estimateEffect(1:6 ~ sotu_type, model.stm, meta = out$meta)
plot(model.stm.ee, "sotu_type", method="difference", cov.value1="speech", cov.value2="written")
```


Topic 1 (Covariate
Level speech Compared to
written)
Topic 4 (Covariate
Level speech Compared to
written)
Topic 5 (Covariate
Level speech Compared to
written)
−0.5
0.0
0.5